บทที่ 3

วิธีการดำเนินงานวิจัย

1. เครื่องมือที่นำมาใช้

1.1 เครื่องคอมพิวเตอร์

1. หน่วยประมวลผลกลาง: Intel T2050 1.6 GHz

2. หน่วยความจำ: 2 GB

3. ระบบปฏิบัติการ: Windows XP SP3

4. โปรแกรมที่ใช้ทดสอบ: Matlab ver. 7.7.0.741

1.2 ชุดข้อมูลที่นำมาใช้ทดลอง

1. ชุดข้อมูล Ruspini ประกอบไปด้วย 75 ตัว มี 2 คุณลักษณะและแบ่งได้ 4 กลุ่ม

2. ชุดข้อมูล Ruspini และบวกกับฟังก์ชั่นเกาส์เซี่ยนเป็นระยะ ๆ เพื่อให้ค่าของข้อมูลไม่คงที่ แต่ยังคงรูปแบบการกระจายที่เหมือนเดิม

ฟังก์ชั่นเกาส์เซี่ยน

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}$$

กำหนดให้ $\mu = 0.1$ และ $\sigma = 1$

3. ชุดข้อมูลอื่น ๆ ซึ่งหาได้จาก ftp://ftp.ics.uci.edu/pub/machine-learning-databases/ ประกอบด้วยรายละเอียดดังนี้

ตารางที่ 3

ตารางแสดงรายละเอียดของชุดข้อมูลที่นำมาทดสอบ

ชื่อชุดข้อมูล	จำนวนกลุ่ม	คุณลักษณะ	จำนวนข้อมูล
Cancer	2	9	683
Iris	3	4	150
Vowel	6	3	871
Wine	3	13	178

1.3 Fitness function

ประเมินระยะทางระหว่างจุดกึ่งกลางของกลุ่มข้อมูลกับข้อมูลแต่ละตัวด้วยระยะทาง Euclidian ตามสูตรดังนี้

$$d(X_n,Z_i) = ||X_n - Z_i||$$

2. อัลกอริทึมของ MSA-PSO Clustering

2.1 กำหนดค่าเริ่มต้นให้แต่ละอนุภาค

1. กำหนดค่าเริ่มต้นให้แก่

$V_i(t), X_i(t), V_{max}, W, C_1$ และ C_2

2. กำหนดค่าเริ่มต้นของขนาดกลุ่มเคลื่อนที่และกำหนดรุ่นหรือแบ่งรุ่น

3. กำหนดค่าเริ่มต้นให้อนุภาคเป็นข้อมูลนำเข้า

2.2 การเวียนเกิดของรุ่น

1. การเวียนเกิดของกลุ่มเคลื่อนที่

2. ปรับค่าของเส้นทาง (velocity) และตำแหน่ง (position) ด้วยสูตร

$$V_i(t) = W^* V_i(t\text{-}1) + C_1 * \text{rand1id} * [X_{pbest}\text{-}X_i(t)] + C_2 * \text{rand2id} * [X_{lbest}\text{-}X_i(t)]$$

2.3 การเปลี่ยนกลุ่มเมื่อถึงระยะเวลาที่กำหนด

2.4 กลุ่มอนุภาคจะคำนวณหาตำแหน่งทั้งหมดใหม่เมื่อเปลี่ยนกลุ่ม

2.5 เลือกกลุ่มเคลื่อนที่ย่อยของอนุภาคที่ได้ค่าที่ค้นหาดีที่สุดในขณะนั้นเพื่อมาผ่านกระบวนการเร่ง ซึ่งอาจจะทำให้กลุ่มเคลื่อนที่ย่อยสามารถค้นหาค่าให้ได้ค่าที่ดีกว่าเดิม จากอัลกอริทึมเดิมของ MSA-PSO ซึ่งอธิบายถึงกระบวนการเร่งแต่ขาดการกำหนดค่าให้กับพารามิเตอร์อีก 2 ตัว คือ r_1 และ δ_i ทำให้ต้องหากระบวนการเร่งใหม่ โดยกำหนดจุดกึ่งกลางที่คาดว่าน่าจะเป็นจุดที่ดี แล้วกำหนดให้กลุ่มตัวเร่งที่เลือกขึ้นมาไปเริ่มค้นหาจากจุดนั้นต่อไป เพื่ออาจจะได้ค่าจุดกึ่งกลางใหม่ที่ดีกว่าจากกลุ่มเคลื่อนที่ย่อยของอนุภาคนั้น ๆ

2.6 ออกด้วยการกำหนดตัวเลขของรุ่น หรือพบข้อกำหนดให้หยุด

3. การกำหนดค่าพารามิเตอร์ให้แต่ละอัลกอริทึม

จากที่มีอัลกอริทึมที่นำมาทดสอบ 3 ตัว และแต่ละอัลกอริทึม จะต้องกำหนดค่าเริ่มต้นหรือพารามิเตอร์เริ่มต้นของตนเอง การค้นหาค่าเริ่มต้นที่เหมาะสมเพื่อให้อัลกอริทึมนั้น ๆ มีประสิทธิภาพในการจัดกลุ่มข้อมูล ใช้วิธีการดังนี้

1. กำหนดให้ทุกอัลกอริทึมเริ่มต้นด้วยจำนวนอนุภาคที่เท่ากัน คือ 30 อนุภาค (สำหรับข้อมูล Ruspini) และมีรอบการเวียนเกิด 1,000 รอบ

2. กำหนดชุดพารามิเตอร์ที่ผ่านการพิจารณาจากงานวิจัยอื่น ๆ มาทั้งหมด 21 ชุด

3. พิจารณาผลของการกำหนดค่าพารามิเตอร์จากข้อ 1 และ 2 ในแต่ละอัลกอริทึม โดยทดสอบกับการจัดกลุ่มข้อมูล Ruspini ซ้ำ 10 ครั้ง เพื่อหาค่าเฉลี่ยของระยะห่างภายในกลุ่มแต่ละกลุ่ม

4. เลือกค่าพารามิเตอร์ที่เหมาะสมมาเพิ่มเติมเป็นชุดพารามิเตอร์ใหม่เพิ่มจาก 21 ชุด แล้วทำการทดสอบกับการจัดกลุ่มข้อมูล Ruspini ซ้ำ 10 ครั้ง นำค่าระยะห่างเฉลี่ยภายในกลุ่มแต่ละกลุ่มที่ได้จากชุดพารามิเตอร์ใหม่มาพิจารณาเลือกค่าที่น้อยที่สุดร่วมกันกับพารามิเตอร์ 21 ชุดเดิม ดังแสดงในตารางที่ 4

5. เมื่อได้ค่าที่น้อยที่สุดของค่าระยะห่างเฉลี่ยภายในกลุ่มข้อมูลแต่ละกลุ่มแล้วนำมาทดสอบกับอัลกอริทึมทั้ง 3 ตัวเพื่อหารอบการเวียนเกิดที่เหมาะสม ทำซ้ำ 20 ครั้ง เพื่อหารอบการเวียนเกิดที่เหมาะสมเพื่อทดสอบประสิทธิภาพในการจัดกลุ่มข้อมูล Ruspini ดังแสดงในตารางที่ 5

ตารางที่ 4

ตารางเปรียบเทียบระยะห่างเฉลี่ยของกลุ่มข้อมูลภายในระหว่าง MSA-PSO กับ PSO-Clustering และ EPSO-Clustering จากการกำหนดค่าพารามิเตอร์เริ่มต้นให้กับแต่ละอัลกอริทึมสำหรับการจัดกลุ่มข้อมูล Ruspini ชนิดคงที่ (ทำซ้ำ 10 ครั้ง ด้วยรอบการเวียนเกิด 1,000 รอบ)

ค่าระยะห่างเฉลี่ยภายในกลุ่มข้อมูล			**ค่าพารามิเตอร์**			
PSO	EPSO	MSA-PSO	C_1	C_2	W	V_{max}
11.6987	13.3388	12.7305	1.4962	1.4962	0.7298	$0.14*R^1$
12.2735	11.5282	24.1177	1.4962	1.4962	0.7298	0.069
12.0296	11.7971	22.7423	1.4962	1.4962	0.7298	0.001

ตารางที่ 4 (ต่อ)

ค่าระยะห่างเฉลี่ยภายในกลุ่มข้อมูล			ค่าพารามิเตอร์			
PSO	EPSO	MSA-PSO	C_1	C_2	W	V_{max}
Error	11.5931	Error	0.005	0.054	0.99	0.069
Error	11.8429	25.3870	0.005	0.054	0.99	0.001
Error	Error	22.4032	0.005	0.054	0.99	$0.14*R^1$
Error	11.5315	24.7102	0.1	2.05	0.95	0.069
Error	11.8035	23.8617	0.1	2.05	0.95	0.001
Error	Error	12.7115	0.1	2.05	0.95	$0.14*R^1$
11.6735	12.1777	11.9938	2	2	0.69	$0.14*R^1$
13.4266	**11.5220**	25.0842	2	2	0.95	0.069
13.4936	11.8171	25.5172	2	2	0.95	0.001
13.5590	13.5572	12.0913	2	2	0.95	$0.14*R^1$
11.9740	11.5310	22.6439	1.49	1.49	0.72	0.069
13.0441	11.8088	24.8541	1.49	1.49	0.72	0.001
12.4140	11.6873	13.0464	1.49	1.49	0.72	$0.14*R^1$
n/a	11.5325	n/a	2	2	0.99	0.069
n/a	11.5266	n/a	1.49	1.49	0.95	0.069
n/a	11.5316	n/a	1.4962	1.4962	0.95	0.069
n/a	n/a	12.6962	2	2	0.72	$0.14*R^1$
n/a	n/a	12.6422	1.49	1.49	0.95	$0.14*R^1$
n/a	n/a	**11.7355**	1.4962	1.4962	0.99	$0.14*R^1$
n/a	n/a	13.5393	1.49	1.49	0.69	$0.14*R^1$

หมายเหตุ: R^1 คือ ค่ารัศมีของชุดข้อมูลคำนวณได้จาก ค่ามากที่สุด-ค่าน้อยที่สุด/2

ตารางที่ 5

ตารางเปรียบเทียบระยะห่างเฉลี่ยของกลุ่มข้อมูลภายในระหว่าง MSA-PSO Clustering PSO-Clustering และ EPSO-Clustering กับการกำหนดรอบการเวียนเกิดให้แต่ละอัลกอริทึมสำหรับการจัดกลุ่มข้อมูล Ruspini ชนิดคงที่ (ทำซ้ำ 20 ครั้ง)

รอบการเวียนเกิด Max_FEs	**ค่าเฉลี่ยระยะห่างภายในของกลุ่มข้อมูล**		
	PSO	EPSO	MSA-PSO
3,000	12.1204	**11.5234**	12.4450
3,500	**11.8153**	11.5261	11.5456
4,000	11.8788	11.5262	n/a
9,000	11.9426	11.5351	**11.5103**

เมื่อพิจารณาข้อมูลที่ได้ในตารางที่ 4 และตารางที่ 5 แล้วจึงกำหนดให้แต่ละอัลกอริทึมมีค่าพารามิเตอร์เริ่มต้นดังนี้

ตารางที่ 6

ค่าพารามิเตอร์เริ่มต้นสำหรับแต่ละอัลกอริทึมที่ใช้ทดสอบ

อัลกอริทึม	C_1	C_2	W	V_{max}	Max_FEs
PSO-Clustering	2	2	0.69	0.069	3,500
EPSO-Clustering	2	2	0.95	0.069	3,000
MSA-PSO-Clustering	1.4962	1.4962	0.99	$0.14*R^1$	9,000

6. เมื่อได้ทำการทดสอบกับชุดข้อมูล Ruspini ชนิดคงที่ แล้วนำไปทดสอบกับข้อมูล Ruspini ชนิดไม่คงที่ และชุดข้อมูลอื่น ๆ ที่นำมาทดสอบร่วม รวบรวมผลการทดสอบของอัลกอริทึมทั้งสามและรายงานผลต่อไป